

a database monitor for confidentiality, quality and tracking of genomic annotations

Valentin Guignon^{1, 2}, Gaëtan Droc¹, Claire Poiron^{1, 3}, Juliette Lengelle¹, Olivier Garsmeur¹, Franc-Christophe Baurens¹ and Stéphanie Bocs¹

Contact: valentin.guignon@cirad.fr

1. UMR DAP, CIRAD, Montpellier, FRANCE

2. Bioversity-France, Bioversity International, Montpellier, FRANCE

3. IMGT IGH, CNRS, Montpellier, FRANCE

1. Context

- Part of GNPAnnot project¹
 - plants, insects, fungi Community Annotation System (CAS)
- Integrated to GMOD² framework:
 - Database: Chado³ (PostgreSQL)
 - Visualisation: GBrowse⁴ (Perl)
 - Editors: Artemis⁵/Apollo⁶ (Java)

- Specific needs:
 - Feature confidentiality → Access Restriction
 - Manual annotation quality → Annotation Inspector
 - Manual annotation tracking → Annotation History

2. Access Restriction

- users and groups handled
- login/password management (with PostgreSQL account sync.)
- scaffold to feature-level access management
- forbidden/read/write access levels supported

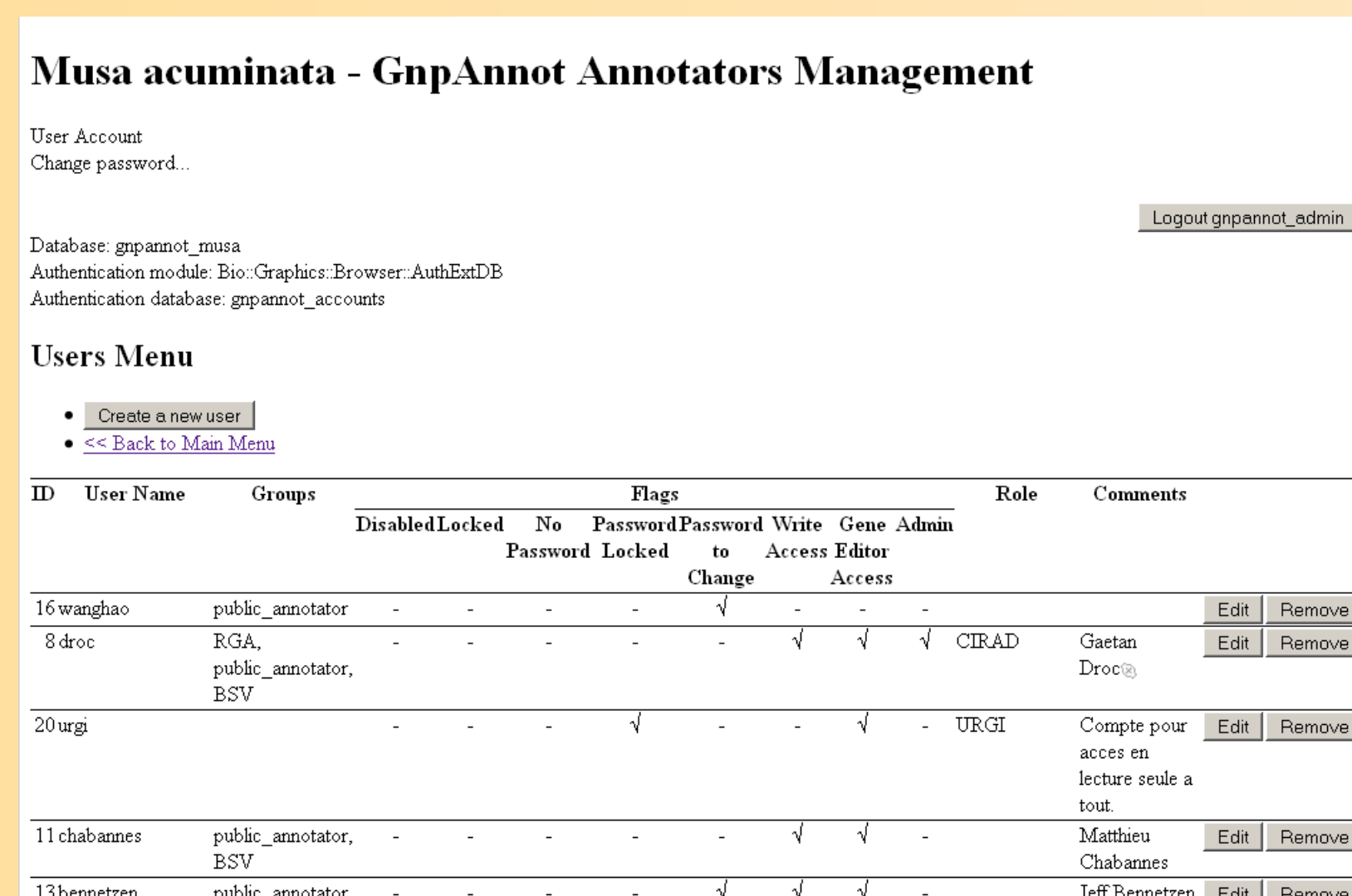


Figure 2. Access Restriction administration interface - user management.

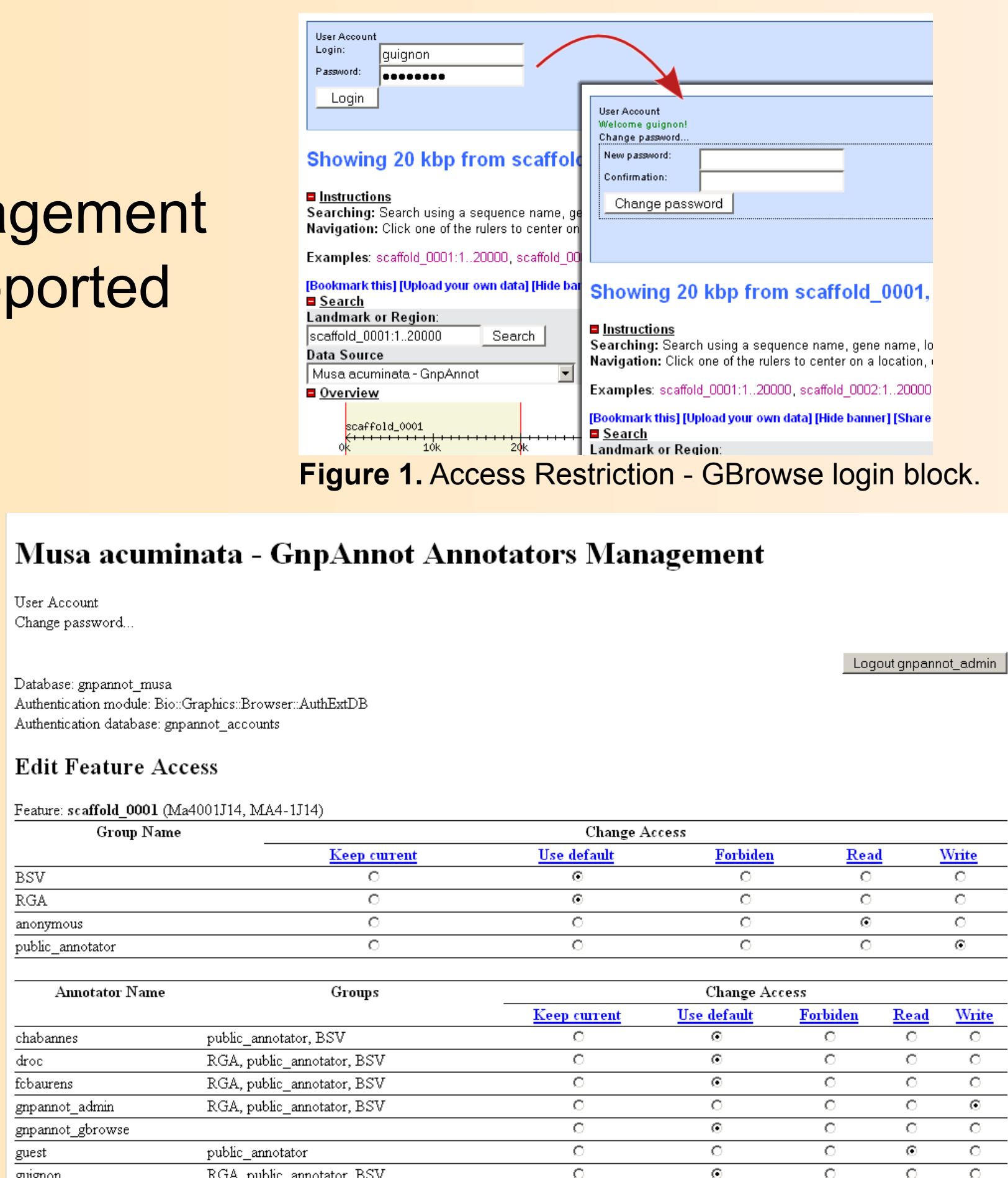


Figure 3. Access Restriction administration interface - contig management.

3. Manual Annotation Inspector

- automated procedures (auto-set qualifiers, transposable element structure)
- validation procedures (structure, sequence content & length, introns, qualifiers and mandatory fields)
- generalisation of Controlled Vocabulary

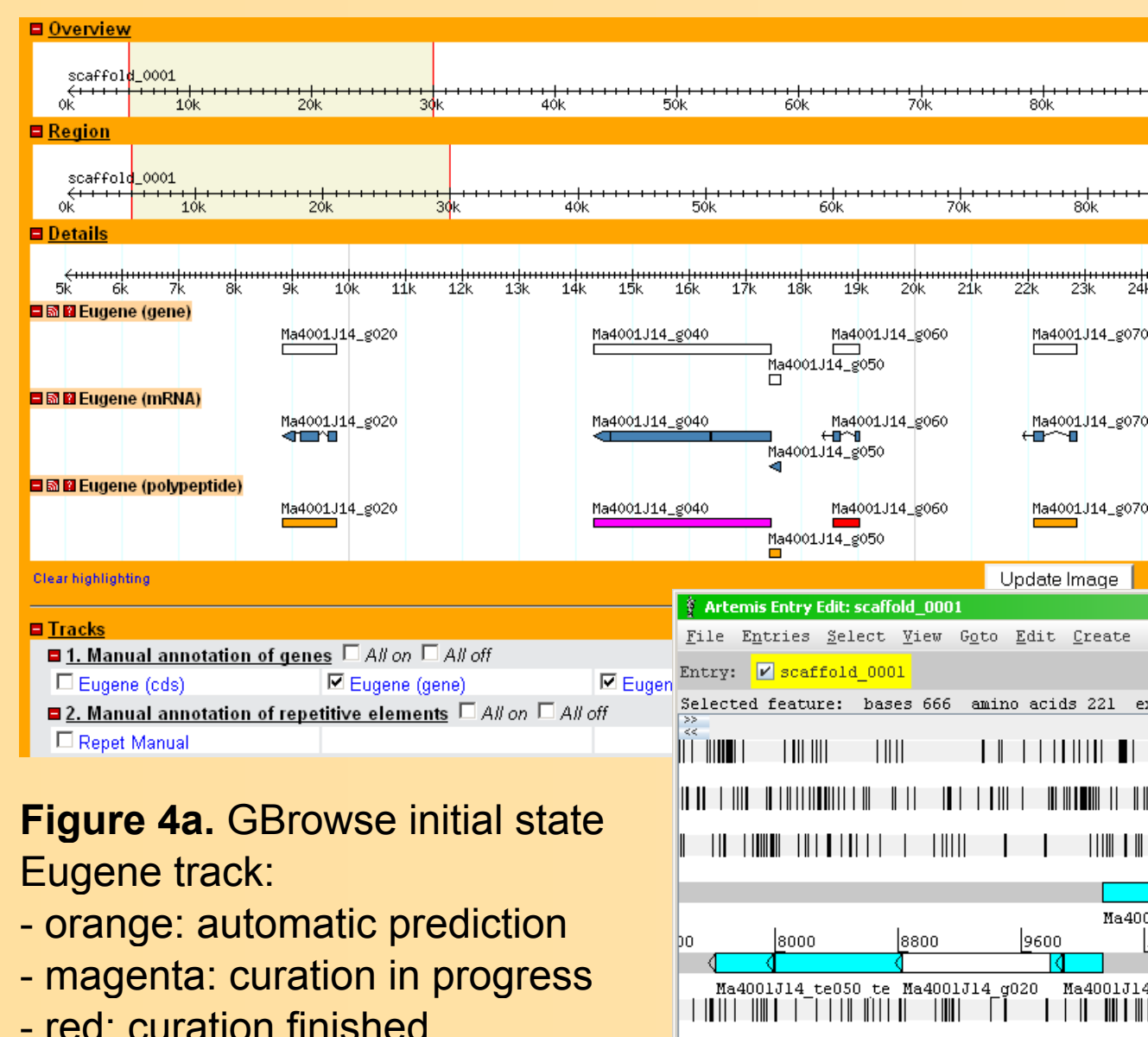


Figure 4a. GBrowse initial state Eugene track:
- orange: automatic prediction
- magenta: curation in progress
- red: curation finished

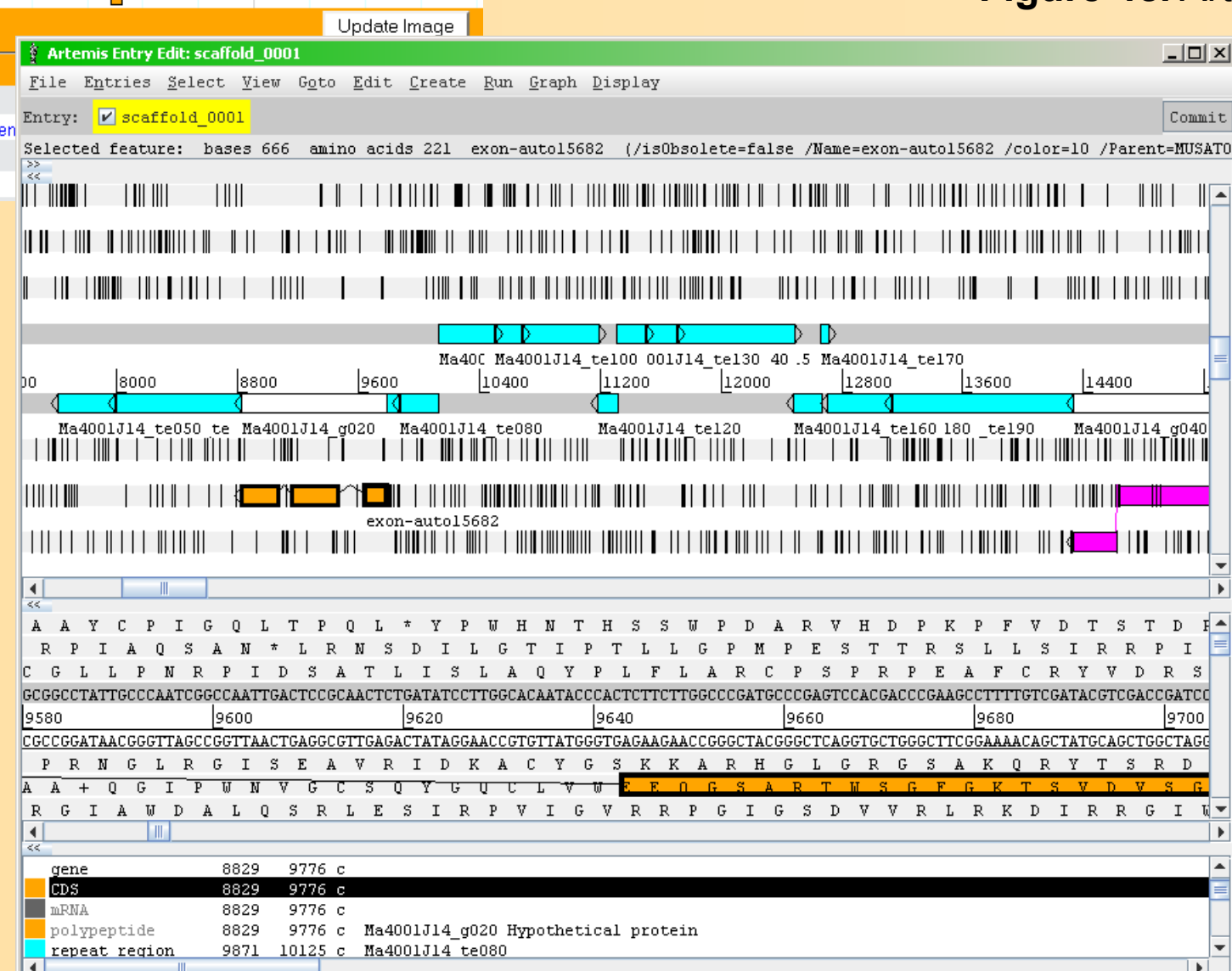


Figure 4b. Artemis: start editing gene Ma4001J14_g020.

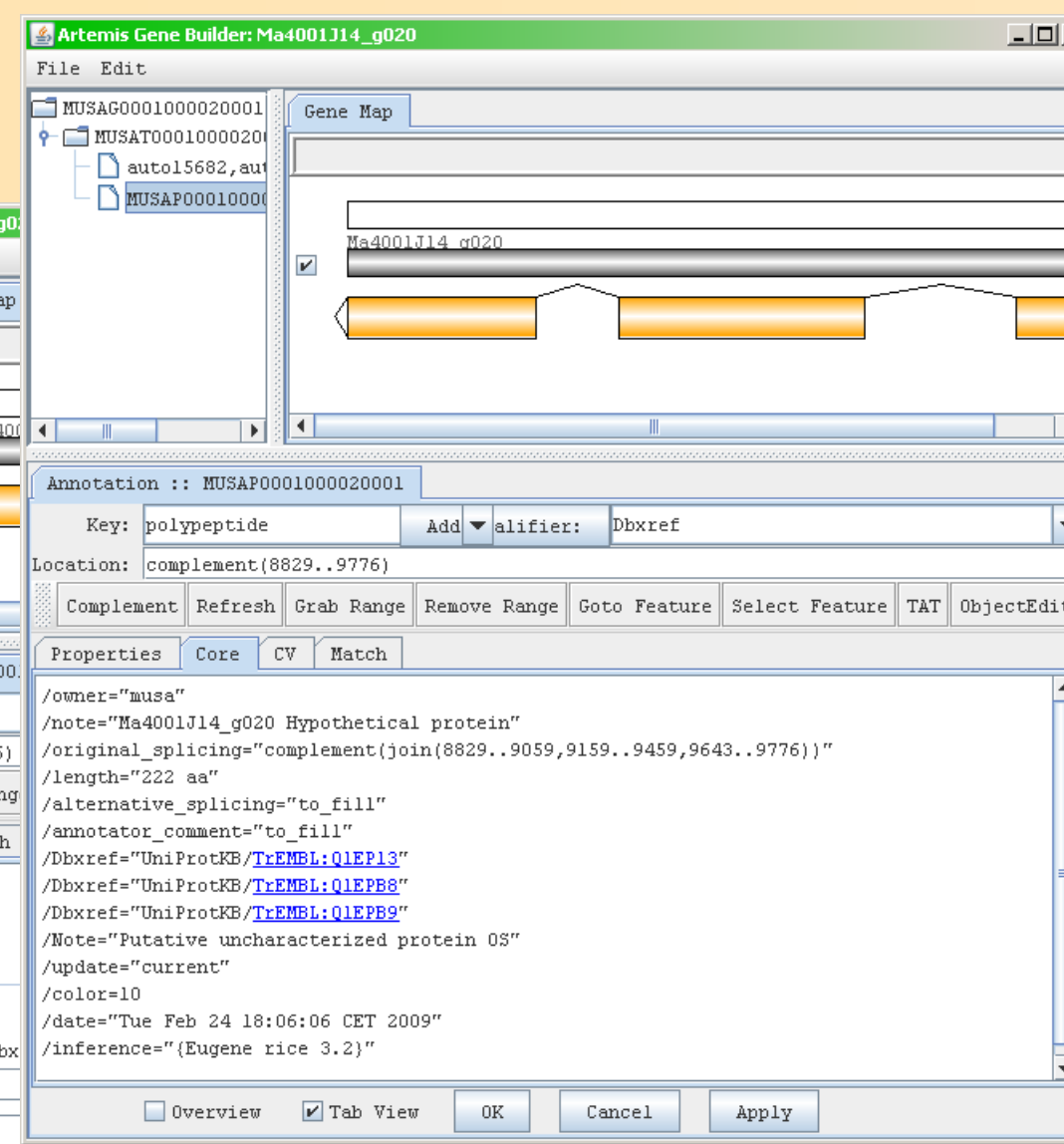


Figure 4c. Artemis Gene Builder: gene Ma4001J14_g020 initial status.

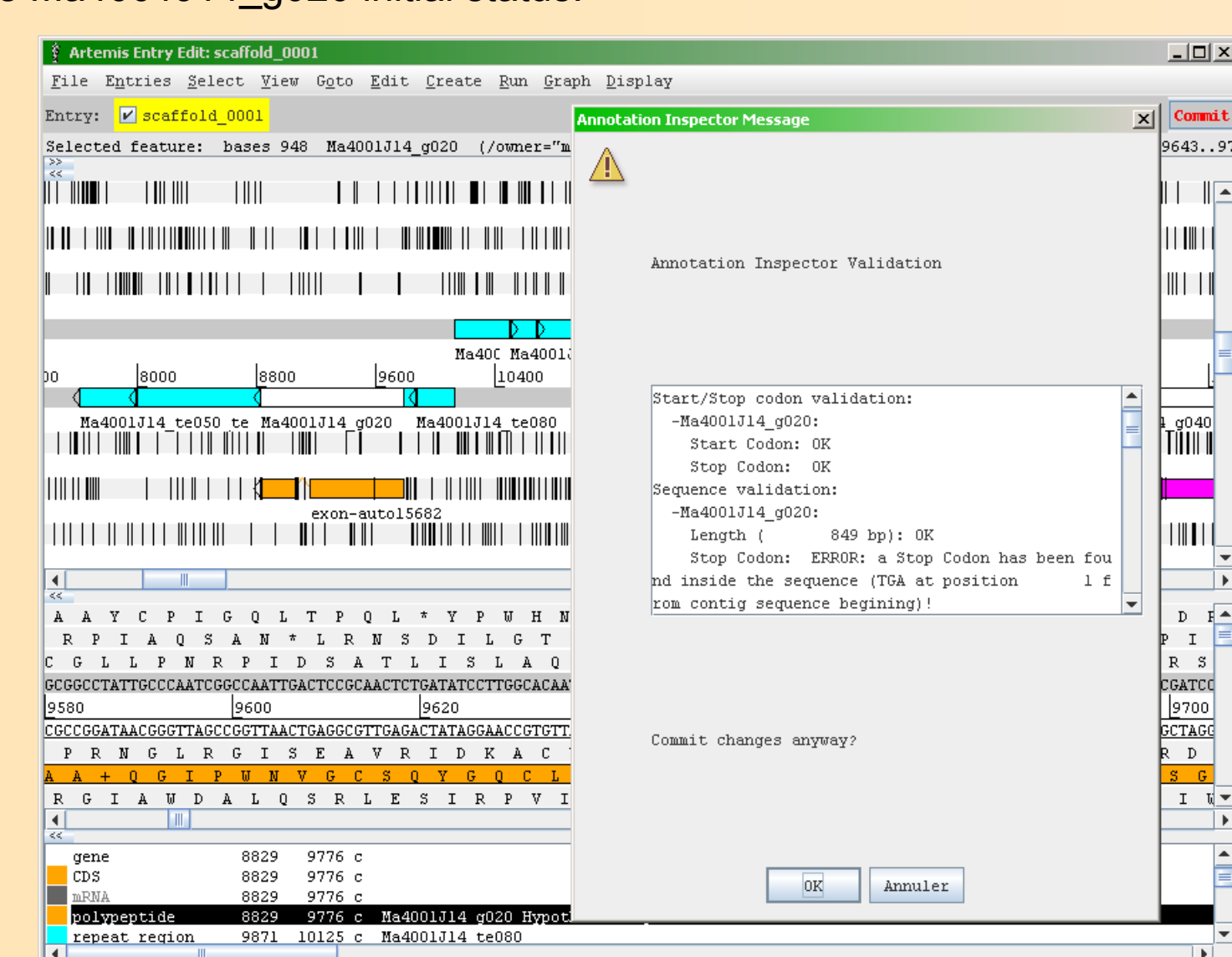


Figure 4d. Artemis: Annotation Inspector message on a commit.

4. Annotation History

- keep track of any modification
- feature history report
- new perspectives (Inspector, Chado undo,...)



Figure 5a. GBrowse History: gene 'Ma4001J14_g020' before curation.



Figure 5b. GBrowse History: gene 'Ma4001J14_g020' after curation.

5. Integration & Compatibility

- GBrowse 1.70⁴
- Artemis⁵
- Compatibility mode (GMOD scripts,...)

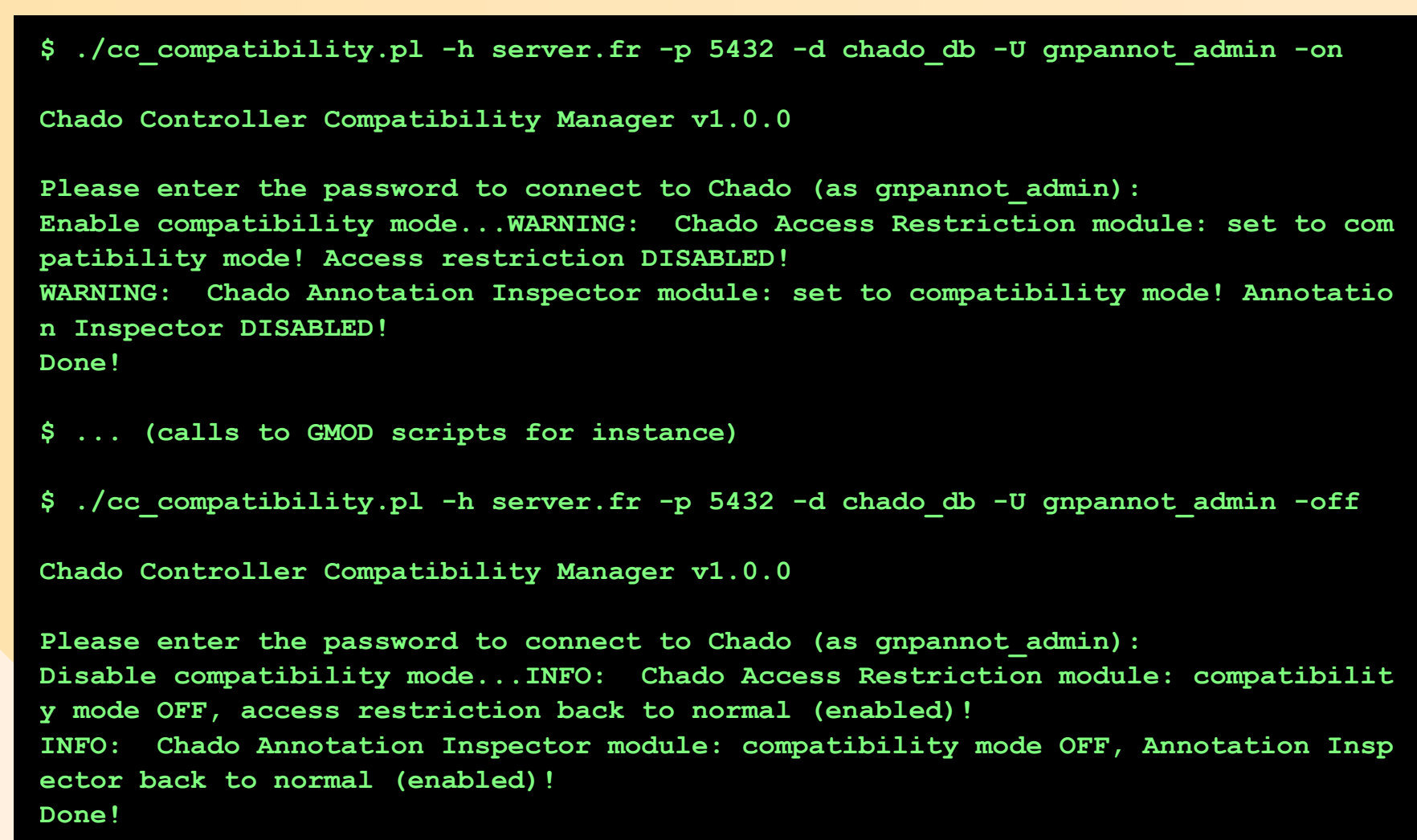


Figure 6. Disabling the Chado Controller in order to turn demanding scripts.
Note: Annotation History can not be turned off.

6. Chado Controller in action!

- Currently used on more than 5 CAS
- 22 Mb of annotated genomic sequences
- 1982 curated genes out of 5004 predicted genes (40%)
- 2703 curated TEs out of 3819 predicted TEs (70%)

7. Perspectives

- GMOD Report⁷ integration
- Tripal⁸ integration
- Apollo⁶ integration
- GBrowse 2⁴ and next generations
- Chado undo/revert script
- Adaptation to other databases

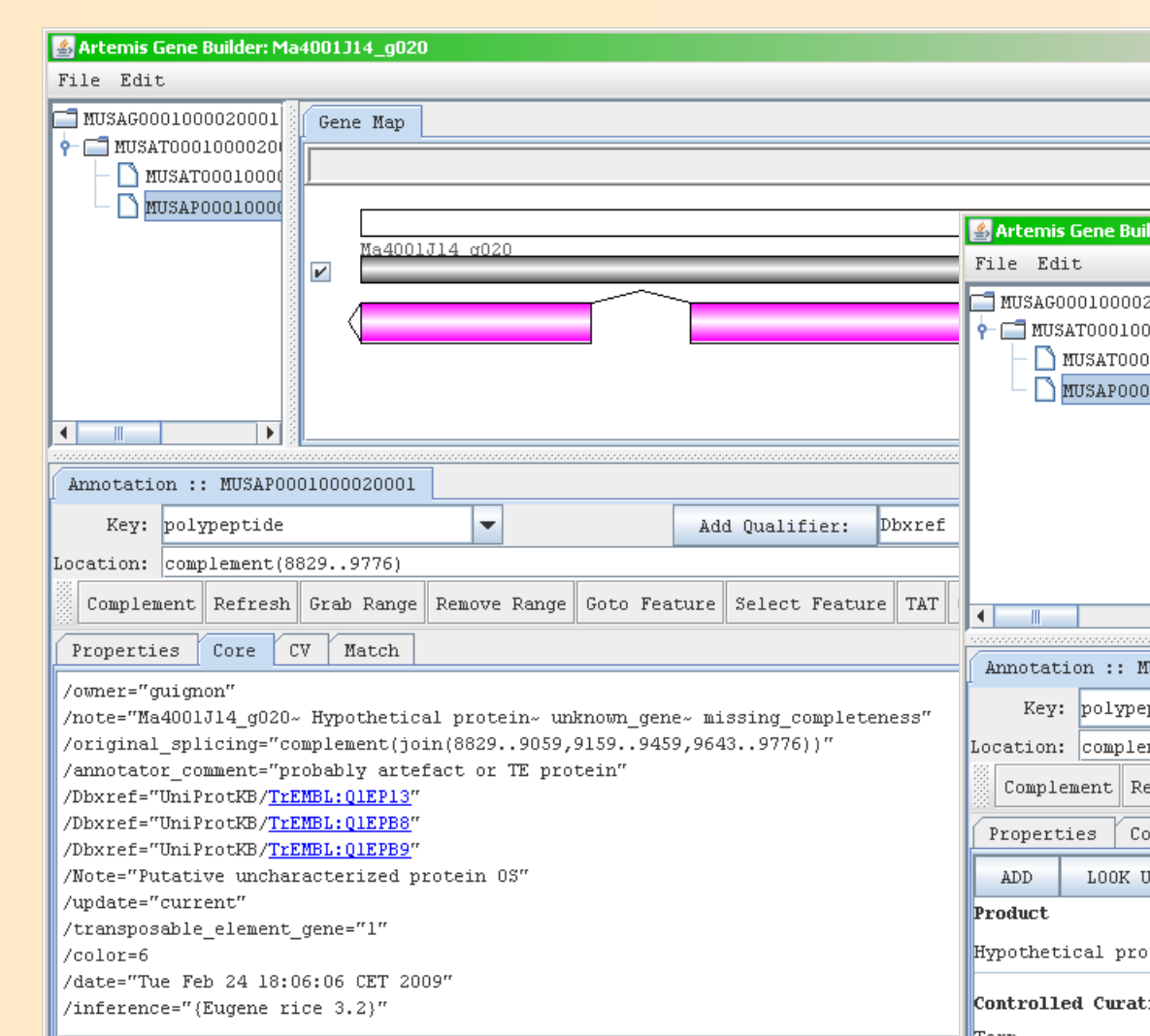


Figure 4f. Artemis Gene Builder: re-opening the gene after the commit.

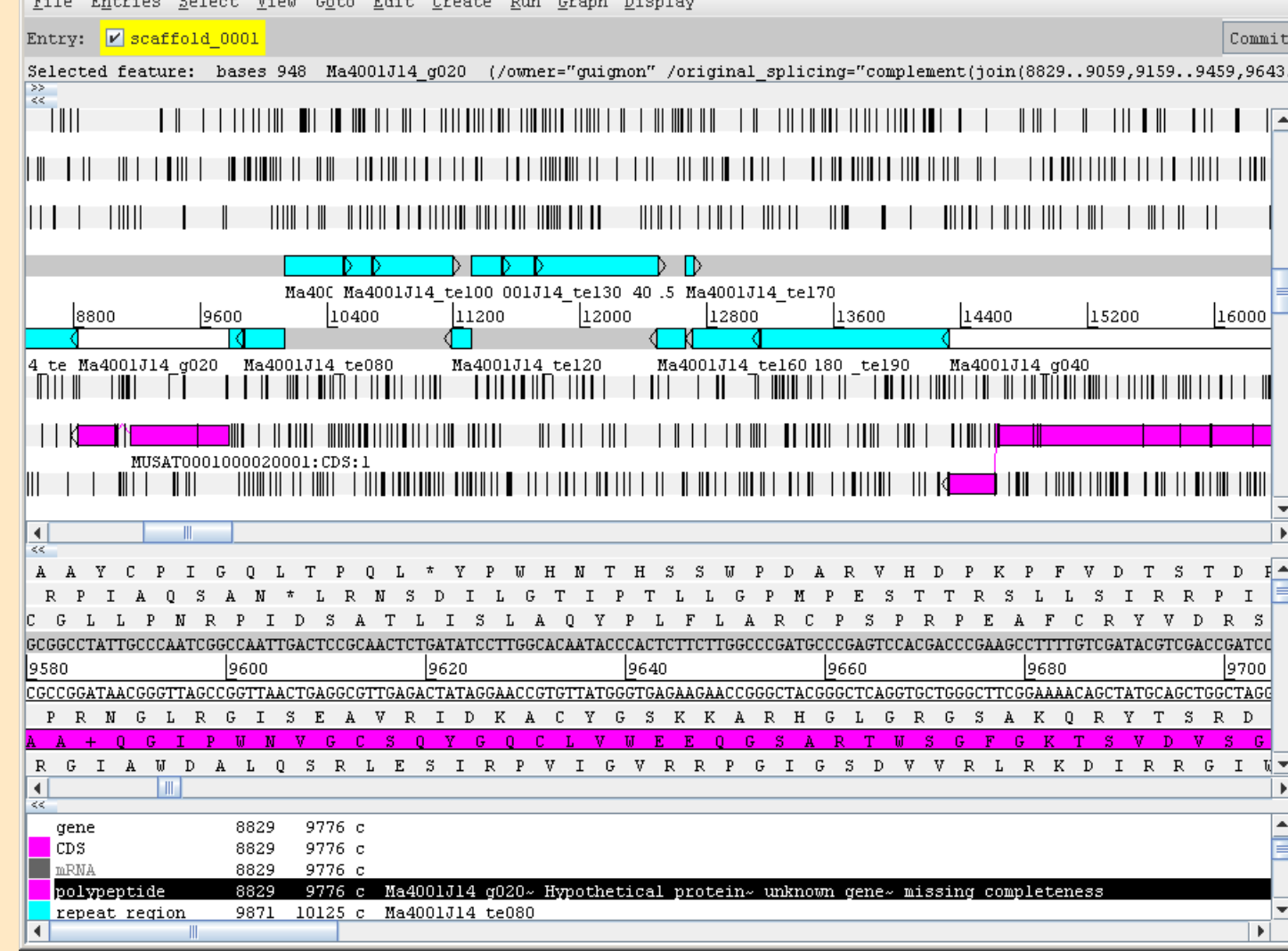


Figure 4e. Artemis: re-opening the region after the commit.

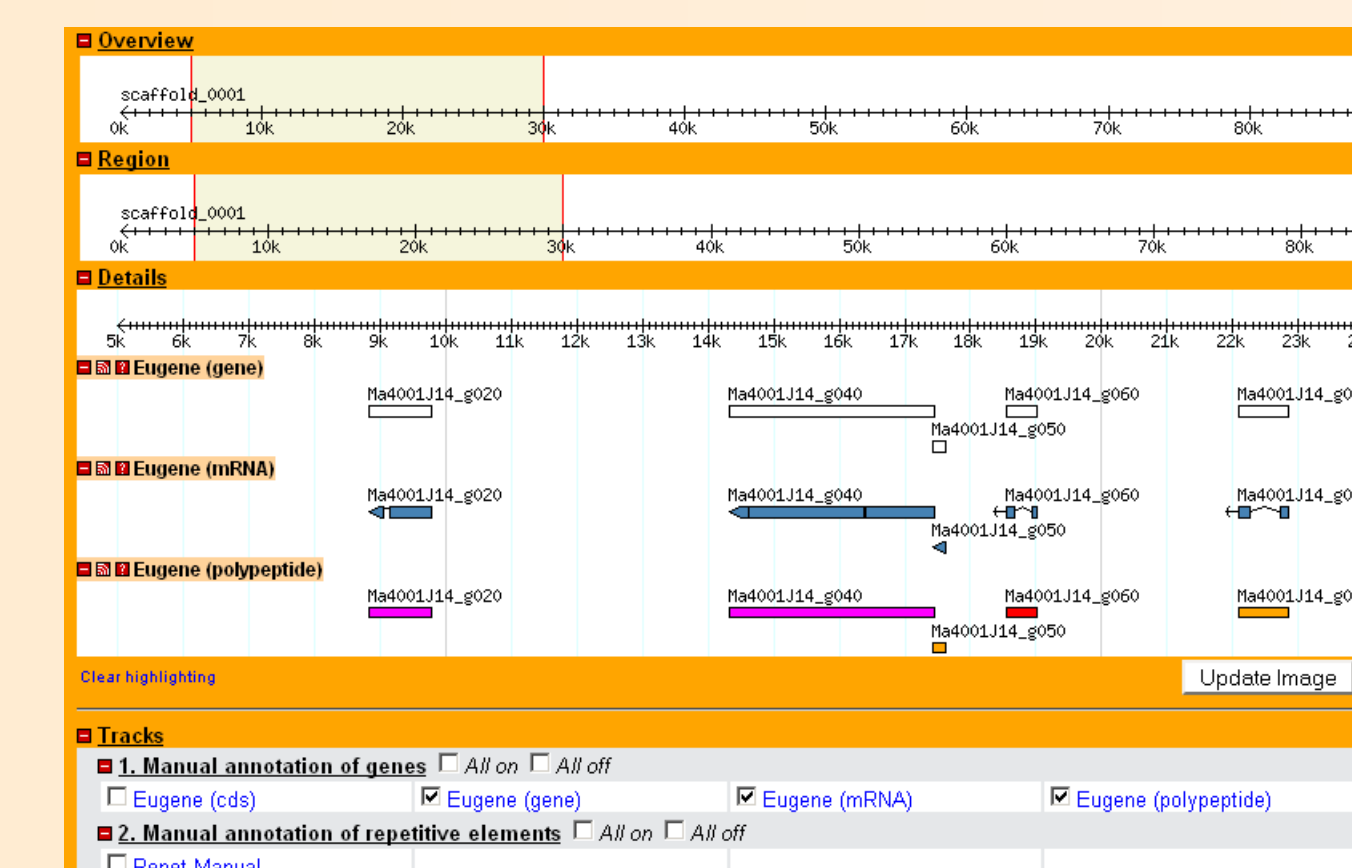


Figure 4g. GBrowse view of the result.

References

- GNPAnnot, <http://www.gnpannot.org>
- GMOD, <http://gmod.org>
- Chado, <http://gmod.org/wiki/Chado>
- Stein LD et al. (2002) The generic genome browser: a building block for a model organism system database. Genome Res 12: 1599-610
- Artemis: sequence visualization and annotation. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA and Barrell B, Bioinformatics (Oxford, England) 2000;16;10;944-5, PUBMED: 11120685
- Apollo: a sequence annotation editor. Lewis SE, Searle SMJ, Harris N, Gibson M, Iyer V, Richter J, Wiel C, Bayraktaroglu L, Birney E, Crosby MA, Kaminker JS, Matthews B, Prochnik SE, Smith CD, Tupy JL, Rubin GM, Misra S, Mungall CJ, Clamp ME (2002). Genome Biology 2002, 3(12):research0082
- GMOD Report, <http://www.aphidbase.com/aphidbase>
- Tripal, <http://www.genome.clemson.edu/software/tripal>

URL du résumé = [http://www.jobim2010.fr/?q=fr/node/55#Chado_Controller: un superviseur pour la confidentialit la qualit et le suivi des annotations](http://www.jobim2010.fr/?q=fr/node/55#Chado_Controller:_un_superviseur_pour_la_confidentialit_la_qualit_et_le_suivi_des_annotat)

The screenshot shows a Windows Internet Explorer browser window. The title bar reads "Posters | JOBIM 2010 - Windows Internet Explorer". The address bar shows the URL: http://www.jobim2010.fr/?q=fr/node/55#Chado_Controller:_un_superviseur_pour_la_confidentialit_la_qualit_et_le_suivi_des_annotat. The browser's menu bar includes "Fichier", "Edition", "Affichage", "Favoris", and "Outils". The toolbar shows "Favoris", "Intranet DIST - Accueil", "Sites suggérés", "Hotmail", "Débuter avec IE8", and "Plus de compléme...". The main content area displays a poster with the following text:

Chado Controller: un superviseur pour la confidentialité, la qualité et le suivi des annotations

Valentin Guignon, Gaëtan Droc, Claire Poiron, Juliette Lengelle, Olivier Garsmeur, Franc-Christophe Baurens and Stéphanie Sidibé Bocs.

Abstract: Dans le cadre du séquençage des génomes, divers systèmes d'annotation ont été développés tel que celui proposé par le projet international Generic Model Organism Database (GMOD). Le projet GNPAnnot a pour but de développer un système d'annotation communautaire (CAS) pour les plantes, les champignons et les insectes basé sur des composants GMOD. Le Chado Controller fait parti des apports du projet GNPAnnot qui complète le coeur du CAS GMOD: base de données Chado, navigateur de génome GBrowse et éditeur Apollo ou Artemis.

Le Chado Controller qui permet de gérer la confidentialité, d'améliorer la qualité des annotations manuelles et de les tracer, comporte trois modules: la restriction des droits d'accès des utilisateurs aux "features", l'inspecteur et l'historique des annotations manuelles. Le module de restriction des droits d'accès permet non seulement de limiter l'accès d'un utilisateur à tout ou partie des données mais également de choisir son niveau d'accès (visualisation seulement ou modification). Pour améliorer la qualité des annotations, l'inspecteur emploie du vocabulaire contrôlé et facilite le travail de l'annotation manuelle en contrôlant le travail du curateur. Il rapporte les incohérence des annotations par rapport à des règles (e.g. structure incorrecte, propriété non ou mal renseignée) et met à jour automatiquement certains champs suite à l'enregistrement des modifications de l'annotateur (e.g. le nom de l'annotateur). Enfin, le module d'historique permet de garder une trace de toutes les modifications apportées aux annotations, quelque soit le type de "feature" (e.g. gène, élément transposable) modifiée et quelque soit l'éditeur utilisé. L'historique des annotations d'une "feature" peut être visualisé à l'aide d'une page Web.

Le Chado Controller, destiné aux bases de données Chado (PostgreSQL), est principalement composé de scripts SQL embarqués dans la base, garantissant ainsi son contrôle global des données. Des parties secondaires du Chado Controller ont été intégrées aux outils GBrowse 1.70 (fenêtre de login, page d'historique) et Artemis (utilisation de l'inspecteur d'annotation). L'intégration de l'inspecteur à un logiciel comme Apollo est aussi simple que d'effectuer des requêtes SQL à partir de celui-ci. Enfin, le Chado Controller est rétro-compatible avec les outils existants et ses différents modules peuvent être activés ou désactivés indépendamment.

Le Chado Controller du projet GNPAnnot est déjà utilisé sur six CAS installé sur la plate-forme South Green (plate-forme de bioinformatique montpelliéraine): sorgho, bananier, canne à sucre, palmier, cacaoyer et caféier. Il a déjà permis l'annotation manuelle de qualité, confidentiellement via le Web, de 964 gènes et de 479 éléments transposables (TE) qui peuvent être facilement exportée dans divers formats à l'aide d'un extracteur.

The Windows taskbar at the bottom shows the "démarrer" button and several open applications: "Reception...", "D:\Mes Do...", "EndNote X3...", "Microsoft E...", "Posters | J...", "Sidibé-Bocs...", "Document1...", and "FR". The system clock shows "11:12".